DOCUMENT RESUME

ED 105 848 IR 001 879

AUTHOR Yens, David P.; Martuza, Victor R.
TITLE Interactive Repeatable Testing: A Many--Faceted Tool.
INSTITUTION Delaware Univ., Newark. School of Education.
PUB DATE Apr 75
NOTE 5p.; Paper presented at the Assocation for Educational Data Systems (Virginia Beach, Virginia, April 29 through May 2, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Attitude Tests; College Students; *Computer Assisted Instruction; Computers; *Data Analysis; Evaluation; Interaction; *On Line Systems; Post Testing; Pretesting; Student Attitudes; *Test Construction; *Testing; Test Interpretation; Time Sharing
IDENTIFIERS *Burroughs B 6700; University of Delaware

ABSTRACT

A computer-based test administration and monitoring system developed at the University of Delaware is described. Using six terminals connected to a Burroughs B-6700 computer, the system provides numerous capabilities for interactive testing and evaluation of students' course learning, while also providing statistical data to faculty in a variety of formats. Student attitudes toward the system have been mixed, although there is a generally positive reaction to the self-pacing features of the system. Faculty have used the data analysis features extensively and have been complimentary of them. Some problems and opportunities for further development of the system are also presented. (DGC)

INTERACTIVE REPEATABLE TESTING: A MANY--FACETED TOOL[1]

David P. Yens and Victor R. Martuza
University of Delaware

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED105848

Many college and public school teachers have, within the last few years, found that the traditional lecture and test method of instruction does not meet the needs of their students for adequate mastery of course materials and interest in course content. This has led to a search for new instructional methods and has resulted in the implementation of open classrooms, independent study, modularized instruction, performance or competency based instruction, and other non-traditional methods. A necessary ingredient for the success of several of these methods is an effective testing program which incorporates the development and administration of test items and the analysis of results. Several people across the country have recently started to use the computer for help in the testing process because of its storage capacity, ability to manipulate large data bases, and analytical capabilities. Most test-related uses of the computer have been oriented toward test scoring and the development and evaluation of large pools of items. For example, tests might be printed by the computer and copied for the students who respond using mark sense forms. These forms are fed into the computer with item identification information and the resulting data may be stored and used for later analyses. An excellent review of several computer assisted test construction projects is provided in the March, 1973 issue of Educational Technology.

Origins of the Computer Assisted Testing Project

At the the University of Delaware, several faculty in the College of Education have developed courses that are partially or completely designed to permit independent study. They were especially concerned with ways of permitting students to demonstrate mastery of each curriculum module or unit. It was believed that the learning process could be improved by permitting a student more than one attempt at a criterion test, but this would require several different copies of a test for each unit. To solve this problem, the faculty turned to the computer.

In 1969 a computer program was developed by Teresa Green to generate randomly parallel tests with a specified number of questions from a large pool of test questions. Test questions were randomly selected so that any item might appear on any copy of the test, but the same item would not appear more than once on a particular test. A blank was provided for the student to enter his response to each question, and a separate strip of the output was reserved for the answers. This answer strip could be separated from the printout containing the questions before the test was given to the student. An instructor could then compare the strip with the student responses and easily grade the test. The randomization of the questions made it possible for a student to take a test two or three times with a low probability of seeing a question repeated. Note that each student had his own copy of the test printout; tests did not need to be reproduced in the usual way.

The basis of this competency-based modular learning process was carefully specified behavioral objectives to guide a student's study of assigned reading materials. A student's grade depended upon the number of points he accumulated for his work. The tests were criterion-referenced in that a pre-specified minimum score had to be achieved before credit could be granted for the unit. Clearly, it was advantageous to permit a student to take more than one test on a unit to insure that he had reached an acceptable level of competence.

Based on our experience, the paper and pencil tests have worked reasonably well, but they have several disadvantages. Several hundred tests must be generated periodically and carried to the testing location. Answer strips must be separated from the questions. Someone must always be present to distribute and score the tests and periodically file them. With large numbers of tests and students these tasks are time-consuming, arduous, and likely to result in errors.

As with all paper and pencil tests, students have the opportunity to view all questions at once and may obtain aid from one question in answering another. The students, of course, may see this as an advantage, but it may lead to large differences in test difficulty so that any given test may not provide a true indication of a student's knowledge. This is especially true if several items refer to one concept. For the courses that have been developed thus far, this problem has not been especially noticeable, but the potential for difficulty exists.

When test items are used repeatedly, it is desirable to obtain item analysis data for the

IR 001 879

[1]Paper given at the 13th Annual Convention of the Association for Educational Data Systems, Virginia Beach, Va. April 29-May 2, 1975

evaluation of the items. In this way, items that are especially easy or difficult or which are confusing or misleading can be detected and rewritten as necessary. With paper and pencil tests, troublesome items can be detected through discussions with students, but the psychometric properties of the items cannot be determined because the random nature and the large number of tests makes the application of item analysis techniques quite difficult.

The Computer Assisted Testing Program

A solution to several of the problems with paper and pencil tests, and the next logical step in the testing program, was the development of an administration process in which the computer administers the test directly to the student. An additional reason for taking this step was the desire to introduce education students to computers before they encounter computer facilities in schools in which they will soon teach. Finally, this was seen as an important step toward the establishment of a comprehensive computer assisted instruction and computer managed instruction program which would meet a critical need within the College.

Three cathode ray tube (CRT) terminals were ordered and programming was initiated in December, 1972. The first students took tests via the terminals in April, 1973. Initially, the program was very simple, but extensive improvements have been made, three more terminals have been obtained, and, recently, upwards of 200 tests a day have been administered.

The six terminals are presently connected via 300 baud acoustic couplers to a general purpose Burroughs B6700 computer that handles all the University's computing. The testing center is open 50 hours a week and is staffed by one full time proctor and several part time student proctors who provide assistance to 800 or so students taking the independent study courses currently on the system. Shortly after the start of a term the students generally know what to do, thus freeing the proctors for additional duties such as data tabulation and maintenance of student files. To provide a backup in the event of occasional computer down time or when the student load becomes too great, paper and pencil tests are available and are administered and scored by the proctors when required.

In our system the students' interaction with the terminal was designed to be as simple as possible. All student responses are set up as formatted input; the student answers each question as it is asked. By faculty request, feedback is provided after each response, and a student continues to respond until he selects the correct answer. A comment option is provided that permits the student to store a narrative comment for each question and for the whole test. These comments are incorporated into the student record file for later review by faculty. Upon completion of the test, a summary of the performance is displayed.

The program contains a number of options for controlling the interaction with students and for generating a variety of student records. Items can be sampled from the item pool with or without replacement for specific students and/ or specific tests. Feedback can be modified or eliminated from the testing process if desired.

The data base contains identification, item response, response latency, and comment information for each test administered. Several options for generating student records are available. Both student-oriented and item-oriented outputs are provided, and several parameters may be manipulated to select on different attributes in the file. This student records system provides a powerful means of analyzing item responses and improving the item pools.

A new addition to the program is an automatic records-keeping routine that automatically stores a student's score in a separate file. Thus, all scores for all tests are directly addressable for each student. This file is dumped periodically to permit students to review their progress and to insure accuracy of the results. Of course, at the end of the semester this file provides a complete record of each student's performance. A special utility routine permits the addition of scores from the paper and pencil tests to this student scores file.

Student and Faculty Attitudes toward CAT

Student attitudes toward the use of computers for testing and instruction have been assessed with a 59-item questionnaire[1] for several terms. Attitude information has also been gained through the use of course attitude surveys and through discussion with the students. In addition, performance data for one course have been analysed.

In contrast to normal uses of attitude surveys, we administered the form both before and after testing and were able to assess changes in attitude. The pre-testing results indicate that students have a slightly positive attitude toward the use of computers for testing and instruction, but are concerned about the mechanical aspects such as the ease of operating a terminal.

The first major use of the attitude questionnaire was during the spring semester of 1974. This turned out to be a period of numerous problems and problem solving activities for our programming staff and the Computing Center staff. System crashes and program difficulties occurred more often than were desirable. Consequently, it was not surprising to find a decrease in positive attitude toward computer uses for testing and instruction. However, on the Likert-type items, the means of the responses were typically rather close to the neutral point which, under the circumstances, indicates that the average student will accept the use of computers for testing even if he encounters occasional computer problems. Furthermore, student attitudes were more positive toward the mechanical aspects of the operation; students found the terminal operation easier than anticipated.

---

(1) This questionnaire is an extensive revision of one developed by Bob Brown at the Pennsylvania State University (Brown, 1966).

Despite the difficulties encountered by the students, an independent course evaluation conducted by the instructor of a psychology course (for which 400 students were taking tests via the terminals) indicated that students strongly preferred the testing by terminals to the more traditional testing process, and they thought they learned more than in a "traditional" course. They viewed the computer quizzing as a very valuable aspect of the course. Almost all students in all courses gave high marks to the opportunity to work at their own pace, knowing the results after each question, and being able to take more than one test. As might be expected, there were some students who intensely disliked the use of terminals or were afraid to use them, no matter what the environment.

During the fall of 1974, the computer and programs worked quite well; the environment was about as good as it will get. A mid-term questionnaire was administered to 130 students. Nineteen percent reported they were afraid to use the terminal or hated to use it. However, thirty percent enjoyed using it, 32 percent didn't mind using it, three percent had no opinion, and 16 percent preferred not to use the terminal. Most of the negative group seemed to be concerned about seeing the whole test and being able to go back and change answers. However, questionnaire results indicated that this problem decreased with increasing terminal use.

Typically, on course evaluation questionnaires that permit students to specify most liked and most disliked features of the course, the use of terminals is mentioned about equally in both categories. This seems to be independent of the ability of the student, his total attitude toward the course, and the testing environment.

The analysis of scores for quizzes taken by the same students on terminals and hard copy indicates that there is no marked difference in performance. It appears that the use of terminals makes little difference even though some students feel that it does.

Faculty have made extensive use of the item analyses provided at the end of each term. Corrections, additions, and deletions to the test question file may be easily made using the time-sharing editor. Faculty who have incorporated the CAT process into their courses have been quite satisfied and have continued to add units and/or questions to their files.

Discussion and Summary

The computer assisted testing model that has been developed has several advantages and some disadvantages for all participating. From the student's point of view, the advantages of this approach are:

(1) The self-pacing capability allows students to work through the course at whatever rate suits them.

(2) The provision of immediate feedback makes the testing activity a learning as well as an evaluation experience.

(3) The multiple testing opportunities provided for each instructional unit permit most students to achieve most of the unit objectives. Anxiety levels are typically lower than for traditional tests; usually, the whole grade does not ride on one or two tests.

(4) The independent study format relieves the student of the task of listening to lectures on content he already knows or could easily learn on his own.

Some of the disadvantages, from the students' viewpoint, relate to the instructional materials used and/or the lack of alternative ways of demonstrating achievement of unit objectives. The self-pacing feature allows students to procrastinate so that there are always a few students who attempt to take all the tests during the last week of class.

From the instructor's point of view, CAT confers at least four advantages:

(1) The level of student achievement is very high. Comprehensive pre- and post measures taken over several semesters have indicated that this method is superior to the standard lecture approach used previously.

(2) The role of the instructor who uses the independent study approach changes considerably. The bulk of his time is now spent in developing and revising curricular materials (both instructional and measurement materials) and tutoring students who experience difficulty with specific concepts or techniques. The latter activity provides a great deal of useful guidance for the curriculum development phase of his work.

(3) If desired, supplementary activities which facilitate transfer of training and the enrichment of student learning can be added on. Such activities may include lectures, field trips, short term internships, research projects, etc. The point is that the CAT model, with independent study, frees the instructor of many highly routinized functions associated with more conventional approaches and allows him to engage in more innovative and profitable activities.

(4) The independent study CAT model allows the instructor to turn an entire course into an ongoing research laboratory in which he can study, in a very controlled fashion, the interrelationships which exist among CAT system variables, content variables, and student variables. The opportunity to do quality and productive instructional research is an invaluable aspect of this approach.

At present, three possible problem areas associated with CAT can be identified:

(1) System reliability is of utmost importance. The implementation of the sophisticated CAT system on a large multipurpose computer attempting to satisfy the diverse and demanding user community resulted at times in an intolerable number of system failures. Clearly, reliability is a necessary condition for the success of any CAT implementation. Fortunately, most of the problems seem to have been worked out and we feel justified in expanding our activity as additional resources become available.

(2) The security of the test item bank is of continuing concern. The CAT approach requires a higher level of security but, by its very nature, it is a very difficult system in which to maintain

security. At present, security is maintained by a maze of user and project numbers, through using the item file in packed form, and through using special keys on the CRT keyboard to permit access to the testing program.

(3) Resource allocation is a continuing problem. Once the success of such a system has been demonstrated, demand grows. The problem is further complicated by different production modes and research demands on the testing. The resolution to this problem depends heavily on the prospects for acquiring additional resources and/or reordering the priorities of the academic units involved.

As mentioned above, outstanding opportunities are afforded by the CAT process for conducting several kinds of research. A few areas for research are related to the following:

(1) Testing time is always at a premium. In CAT courses, students spend a great deal of time testing. Therefore, testing efficiency is of great concern. How can the maximum amount of information regarding student achievement be obtained in the least amount of test time? A number of researchers are currently working on this problem. Some potential answers may lie in sequential (or tailored) testing, in the application of Bayesian techniques to item selection and analysis, and possibly in the use of confidence measures.

(2) The effects of variables such as type of feedback, spacing of testing opportunities, size and nature of the item banks, setting of mastery cut points, etc., on student learning and attitudes have yet to be systematically studied.

(3) At this time, we know very little about the ways in which students use the CAT system. A catalogue of such strategies, an evaluation of their relative effectiveness, and the determination of whether or not they can be successfully used by others is clearly appropriate and important at this time.

Computer assisted testing has unquestionably expanded the repertoire of instructional strategies available to faculty. However, the full range of the effects of such strategies on student learning of and attitudes toward course content have yet to be assessed. CAT appears to have great potential; a systematic exploration of its usefulness is now in order.

BIBLIOGRAPHY

Brown, B. R., An instrument for the measurement of expressed attitude toward computer-assisted instruction. In Semi-Annual Progress Report No. R-5, Experimentation with Computer-Assisted Instruction, prepared by Harold E. Mitzel et.al., University Park: The Pennsylvania State University, August 8, 1966, 35-37.